Amendments to the Specification:

Please replace the paragraphs at page 46, lines 15-28 with the version that follows below:

Associations of chromosomal localizations for mapped genes with amplicons implicated in cancer are based on literature searches (PubMed http~~://www. ncbi. nlm. nih. gov/entrez/query. fcgi~~), OMIM searches (Online Mendelian Inheritance in Man~~, http ://www. ncbi. nlm. nih. gov/Omim/searchomim. html~~) and the comprehensive database of cancer amplicons maintained by Knuutila, et al. (Knuutila, et al., DA copy number amplifications in human neoplasms. Review of comparative genomic hybridization studies. Am J Pathol 152: 1107-1123,1998~~. http ://www. helsinki. fi/~lgl www/CMG. html~~). For many of the mapped genes, the cytogenetic region from Knuutila is listed followed by the number of cases with documented amplification and the total number of cases studied.

For single nucleotide polymorphisms, an accession number is given if the SNP is documented in dbSNP (the database of single nucleotide polymorphisms) maintained at NCBI ~~(http ://www. ncbi. nlm. nih. gov/SNP/index. html)~~. None of the sequences used in this application have SNPs represented in dbSNP.

Please replace the paragraph at page 49, lines 13-23 with the version that follows below:

For a number of protein phosphatases of the invention, there is provided a classification of the protein class and family to which it belongs, a summary of noncatalytic protein motifs, as well as a chromosomal location. This information is useful in determing function, regulation and/or therapeutic utility for each of the proteins. Amplification of chromosomal region can be associated with various cancers. For amplicons discussed in this application, the source of information was Knuutila, et al (Knuutila S, Bjorkqvist A-M, Autio K, Tarkkanen M, Wolf M, Monni O, Szymanska J, Larramendy ML, Tapper J, Pere H, El-Rifai W, Hemmer S, Wasenius V-M, Vidgren V & Zhu Y: DNA copy number amplifications in human neoplasms. Review of comparative genomic hybridization studies. Am J Pathol 152: 1107-1123,1998~~. http ://www. helsinki. fi/~lgLwww/CMG. html~~).

Please replace the paragraph at page 51, lines 1-10 with the version that follows below:

SGP061, SEQ ID NO : 2 is a novel MKP-like phosphatase. The dual specificity phosphatase family includes around 20 known human members ~~(for a list, see http://smart.emblheidelberg. de/smart/getmembers. pl? WHAT=species & NAME=DSPc & WHICH=Homo sapiens)~~. Well-known members of the MPK family of dual-specificity phosphatases include: DUS1 (also known as MPK-1, CL100, PTPN-10, erp, VH1 or 3CH134), DUS3 (also known as VHR), DUS4 (also known as HVH2, TYP1, MKP2 or VH2), DUS5 (also known as HVH3, B23, VH3), DUS6 (also known as PYST1, MKP3, rVH6), DUS7 (also known as PYST2), CDKN3 (also known as CDKN3, KAP, CIP2 or CDI1), VH5 and STYX.

Please replace the paragraph at page 107, line 1-18 with the version that follows below:

Table 2 lists the following features of the genes described in this application: chromosomal localization, single nucleotide polymorphisms (SNPs), representation in dbEST, and repeat regions. From left to right the data presented is as follows: "Gene Name","IDiMna","ID#aa","FL/Cat","Superfamily","Group","Family", "Chromosome","SNPs","dbESThits", & "Repeats". The contents of the first 7 columns (i. e.,."Gene Name","ID#na","ID#aa","FL/Cat","Superfamily","Group", "Family") are as described above for Table 1."Chromosome"refers to the cytogenetic localization of the gene. Information in the"SNPs"column describes the nucleic acid position and degenerate nature of candidate single nucleotide polymorphisms (SNPs)."dbEST hits"lists accession numbers of entries in the public database of ESTs (dbEST~~, http ://www. ncbi. nlm. nih. gov/dbEST/index. html~~) that contain at least 100 bp of 100% identity to the corresponding gene. These ESTs were identified by blastn of dbEST."Repeats"contains information about the location of short sequences, approximately 21 bp in length, that are of low complexity and that are present in several distinct genes. These repeats were identified by blastn of the DNA sequence against the non-redundant nucleic acid database at NCBI (nrna). To be included in this repeat column, the sequence typically has 100% identity over its length and is present in at least 5 different genes.

Please replace the paragraph at page 109, lines 7, 12, 20 with the version that follows below:

Table 3 lists the extent and the boundaries of the phosphatase catalytic domains. The column headings are :"Gene Name","ID#na","IDaa","FL/Cat", "Domain", "Phosstart","Phosend","Profllestart","Profileend","Other Domains" and "SH2 Boundaries."The contents columns"Gene Name","ID#na", "ID#aa","FL/Cat",are as described above for Table 1."Phos Start","Phos End", "Profile Start"and"Profile End"refer to data obtained using a Hidden-Markov Model to define catalytic range boundaries ~~(http://pfam. wustl. edu/index. html)~~. The boundaries of the catalytic domains within the overall protein are noted in the"Phos Start"and"Phos End"columns. Three profiles were used, one for dual specificity phosphatases (DSP) which is 173 amino acids long ;, one for STPs, which is 301 amino acids long ; and one for PTPs, which is 264 amino acids long. ~~(The profiles used are described in http ://pfam. wustl. edul)~~. Proteins in which the profile recognizes a full length catalytic domain have a"Profile Start"of 1 and, for the three families, the following Profile Ends: 173 for DSP, 301 for STPs, and 264 for PTPs. Genes which have a partial catalytic domain will have a"Profile Start"of greater than 1 (indicating that the beginning of the phosphatase domain is missing, and/or a"Profile End"of less than 261 (indicating that the C-terminal end of the phosphatase domain is missing). The"Other domains" column lists non-phosphatase domains identified in the novel phosphatase proteins by PFAM searching ~~(http ://pfam. wustl. edu/)~~. SGP057, SEQ ID NO : 1, contains two partial SH2 domains.

Please replace the paragraph at page 111, lines 1-18 with the version that follows below:

Table 4 describes the results of Smith Waterman similarity searches (Matrix: Pas 100; gap open/extension penalties 12/2) of the amino acid sequences against the NCBI database of non-redundant protein sequences ~~(http ://www. ncbi. nlm. nih.gov/Entrez/protein.html)~~. The column headings are: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Family", "Pscorc", "aajcngth", "aaIDmatch", "%Identity", "%SimilaritY", "ACC#nraamatch", "Description" ! The contents of columns,"Gene Name", "ID#na", "ID#aa",

"FL/Cat", and"Family" are as described above for Table 1."Pscore"refers to the Smith Waterman probability score. This number approximates the probability that the alignment occurred by chance. Thus, a very low number, such as 2.10E-64, indicates that there is a very significant match between the query and the database target."aa~length" refers to the length of the protein in amino acids."aaIDmatch"indicates the number of amino acids that were identical in the alignment."% Identity"lists the percent of nucleotides that were identical over the aligned region."% Similarity" lists the percent of amino acids that were similar over the alignment. "ACC#nraa~match" lists the accession number of the most similar protein in the NCBI database of non-redundant proteins."Description"contains the name of the most similar protein in the NCBI database of non-redundant proteins.

Please replace the paragraph at page 113, line 11-29 with the version that follows below:

Novel phosphatases were identified from the Celera human genomic sequence databases, and from the public Human Genome Sequencing project ~~(http://www.ncbi.nlm.nih.gov/)~~ using hidden Markov models (HMMRs). The genomic database entries were translated in six open reading frames and searched against the model using a Timelogic Decypher box with a Field programmable array (FPGA) accelerated version of HMMR2. 1. The DNA sequences encoding the predicted protein sequences aligning to the HMMR profile were extracted from the original genomic database. The nucleic acid sequences were then clustered using the Pangea Clustering tool to eliminated repetitive entries. The putative protein phosphatase sequences were then sequentially run through a series of queries and filters to identify novel protein phosphatase sequences. Specifically, the HMMR identified sequences were searched using BLASTN and BLASTX against a nucleotide and amino acid repository containing known human protein phosphatases and all subsequent new protein phosphatase sequences as they are identified. The output was parsed into a spreadsheet to facilitate elimination of known genes by manual inspection. Two models were developed, a"complete"model and a "partial"or Smith Waterman model. The partial model was used to identify subcatalytic phosphatase domains, whereas the complete model was used to identify complete catalytic domains. The selected hits were then queried using BLASTN against the public nrna and EST databases to confirm they are indeed unique. In some cases the novel

genes were judged to be orthologues of previously identified rodent or vertebrate protein phosphatases.

Please replace the paragraph at page 114, lines 4-15 with the version that follows below:

Extension of partial DNA sequences to encompass the full-length openreading frame was carried out by several methods. Iterative blastn searching of the cDNA databases listed in Table 5 was used to find cDNAs that extended the genomic sequences."LifeGold"databases are from Incyte Genomics, Inc ~~(http://www. incyte. com/)~~. NCBI databases are from the National Center for Biotechnology Information ~~(http://www. ncbi. nlm. nih. gov/)~~. All blastn searches were conducted using a blosum62 matrix, a penalty for a nucleotide mismatch of-3 and reward for a nucleotide match of 1. The gapped blast algorithm is described in: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),"Gapped BLAST and PSI BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25: 3389-3402).

Please replace the paragraph at page 115, line 11-16 with the version that follows below:

Another method involved using the Genewise program ~~(http ://www. sanger. ac. uk/Software/Wise2/)~~ to predict potential ORFs based on homology to the closest orthologue/homologue. Genewise requires two inputs, the homologous protein, and genomic DNA containing the gene of interest. The genomic DNA was identified by blastn searches of Celera and Human Genome Project databases. The orthologs were identified by blastp searches of the NCBI non-redundant protein database (NRAA). Genewise compares the protein sequence to a genomic DNA sequence, allowing for introns and frameshifting errors.

Please replace the paragraph at page 118, lines 2-9 with the version that follows below:

For genes that were extended using Genewise, the accession numbers of the protein ortholog and the genomic DNA are given. (Genewise uses the ortholog to assemble the

coding sequence of the target gene from the genomic sequence). The amino acid sequences for the orthologs were obtained from the NCBI non-redundant database of proteins. ~~(http://www. ncbi. nlm. nih. gov/Entrez/protein. html)~~. The genomic DNA came from two sources: Celera and NCBI-NRNA, as indicated below. cDNA sources are also listed below. Abbreviations: HGP: Human Genome Project; NCBI, National Center for Biotechnology Information.

Please replace the paragraph at page 123, lines 21-29 with the version that follows below:

"cDNA libraries"derived from a variety of sources were immobilized onto nylon membranes and probed with 32P-labeled cDNA fragments derived from the gene (s) of interest. The sources of RNA were: 1) Biochain Institute (Hayward, CA~~; http://www. biochain. com/main~3. html~~) ; 2) Clontech (Palo Alto, CA~~, http ://www.clontech.com/~~) 3) mammalian cell lines used by the National Cancer Institute (NCI) Developmental Therapeutics Program ~~(http ://dtp. nci. nih. gov/; can be ordered from ATCC: http://www. atcc. ora/catalogs. html)~~ ; 4) PathAssociates ~~(http ://www. saic.com/companv/subsidiaries/pai.html ;~~ San Diego, California). The protocols for preparing cDNA arrays are detailed below. Several cell lines were treated with compounds to evaluate their effects on gene expression. There were eight treatments: 1) control, 2) low serum, 3) 200uM mimosine, 4) 3mM HU, 5) 2uM AUR2 inhibitor, 6) lOuM cisplatin, 7) 400 ng/ml nocodozole-24 hours, and 8) 400 ng/ml nocodozole-48 hours.

Please replace the paragraph that beings at page 127, lines 10 and ends at page 128, line 5 with the version that follows below:

Several sources were used to find information about the chromosomal localization of the genes in the present invention. The Celera browser was used to localize celera configurations to specific cytogenic bands ~~(http ://www. celera. com)~~. Also, the accession number for the nucleic acid sequence was used to query the Unigene database. ~~The site containing the Unigene search engine is: http://www. ncbi. nlm. nih. gov/UniGene/Hs. Home. html.~~ Information on map position within the Unigene database is imported from several

sources, including the Online Mendelian Inheritance in Man ~~(OMIM, http ://www. ncbi. nlm. nih. gov/Omim/searchomim. html)~~, The Genome Database ~~(http ://gdb. infobiogen. fr/gdb/simpleSearch. html)~~, and the Whitehead Institute human physical map ~~(http ://carbon. wi. mit. edu: 8000/cgi- bin/contig/sts~inf ? database=release)~~. If Unigene has not mapped the EST, then the nucleic acid for the gene of interest is used as a query against databases, such as dbsts and htgs ~~(described at http://www. ncbi. nlm. nih. gov/BLAST/blastdatabases. html)~~ containing sequences that have been mapped already. The nucleic acid sequence is searched using BLAST-2 at NCBI ~~(http ://www. ncbi. nlm. nih. gov/cgi-bin/BLAST/nph-newblast)~~ and is used to query either dbsts or htgs. Once a cytogenetic region has been identified by one of these approaches, disease association is established by searching OMIM with the cytogenetic location. OMIM maintains a searchable catalog of cytogenetic map locations organized by disease. A thorough search of available literature for the cytogenetic region is also made using Medline ~~(http://www. ncbi. nlm. nih. gov/PubMed/medline. html)~~. References for association of the mapped sites with chromosomal abnormalities found in human cancer can be found in: Knuutila, et al., Am J Pathol, 1998,152: 1107-1123.

Please replace the paragraph at page 129, lines 9-20 with the version that follows below:

The most common variations in human DNA are single nucleotide polymorphisms (SNPs), which occur approximately once every 100 to 300 bases. Because SNPs are expected to facilitate large-scale association genetics studies, there has recently been great interest in SNP discovery and detection. Candidate SNPs for the genes in this patent were identified by blastn searching the nucleic acid sequences against the public database of sequences containing documented SNPs ~~(dbSNP, at NCBI, http ://www. ncbi. nlm. nih. gov/SNP/snpblastpretty. html)~~. dbSNP accession numbers for the SNP-containing sequences are given. SNPs were also identified by comparing several databases of expressed genes (dbEST, NRNA) and genomic sequence (i. e., NRNA) for single basepair mismatches. The results are shown in Table 2, in the column labeled"SNPs". These are candidate SNPs-their actual frequency in the human population was not determined. The code below is standard for representing DNA sequence: